Call for Programmers

Software Design Lab Spring 2019

Rob Garfield

Description

I need a Python module that takes in a large set of retweets (provided outside of the module from a Twitter scraper) and builds two csv files – a nodes and an edges file – for use in creating network visualizations of Twitter interactions.

Requirements

The module will need to be able to take in a csv from a Twitter scraper and convert it into a list of Tweeters (nodes) and their connection to those they retweet (edges) in csv format.

The input csv can be quite large, so the module will need to be able to handle large amounts of text data in a reasonable time frame.

The module will need to provide an API (can just be a class object or even just functions) that is usable in any other Python app in Python 3.6 or higher.

The specification of the input csv is negotiable.

To avoid portability issues, the module must not use outside libraries.

The module must provide documentation both within the code and outside of it as to its proper use.

The module must be publicly available for download, forking, and use from GitHub; and must include a free software license.

Why is this cool?

This will help all Digital Humanities and/or Social Media researchers cut time off the process of conducting Twitter-based social research. The output can be used to generate network analyses in a myriad of visualization tools. It will be put to immediate use in developing a tracking system for the interaction of contemporary poets on social media.